Development of semantically rich retrofit 3D models

Farhad Sadeghineko[1] and Bimal Kumar[2]

[1]Department of Construction and Surveying, Glasgow Caledonian University, Glasgow, United Kingdom, Email: farhad.sadeghineko@gcu.ac.uk

[2]Department of Architecture and Built Environment, Northumbria University, Newcastle upon Tyne, United Kingdom, Email: bimal.kumar@northumbria.ac.uk

Abstract

The use of Building Information Modelling (BIM) has gained considerable interest in new build projects. However, its use in existing assets has been limited to geometric models utilising Point Cloud Data (PCD) as the primary source of data. The inclusion of non-geometrical data from distributed sources in the geometric model to make it semantically rich has been fraught with considerable challenges. In this paper, an approach is proposed to provide a framework for generating semantically-rich parametric models for existing assets. While the geometric information like length, width, area, and volume can be extracted from a PCD, non-geometric data may need to be appended to this for generating genuinely semantically rich models. The Comma Separated Values (CSV) format is utilised to represent the data that can be extracted from PCDs. In addition, the non-geometric information derived from other sources is appended to the CSV file. Subsequently, the Resource Description Framework (RDF) data is generated from the data presented in the CSV files. RDF is a commonly used Semantic Web technology for storing, sharing, and reusing information on the Web. The RDF data is then used to create the IFC data model by translating RDF into IFC. The IFC file is used to generate 3D BIM by importing it into any IFC-compliant application. The proposed approach was validated on one part of the Edinburgh castle, a relatively complex historical building. The choice of building for validating the approach was driven by technical as well as pragmatic reasons. Technically, the approach will have proven its robustness if it could be shown to work for a complex rather than a relatively simple building. Pragmatically, the authors had access to data on Edinburgh Castle due to an ongoing partnership with the Historic Environment Scotland (HES). However, as a result of the validation process, it is suggested that the proposed approach should be applicable to any existing building.

1 Introduction and Background

While Building Information Modelling (BIM) process has recently gained a lot of momentum in new build projects in Architecture, Engineering, and Construction (AEC) for varying purposes like design, construction as well as facility management, its use in existing buildings has been hampered by the challenges and limitations of involved technologies [136-099-042]. In recent years, 3D laser scanning technology, as a remote sensing technique, has been extensively used to collect geometrical data from existing buildings. The output of this technology is a set of three-dimensional point measurements, also known as Point Cloud Data (PCD). Several approaches have been proposed employing PCD as the primary geometric data source to map building models. However, in order to transform such a model containing only geometrical data into a 3D parametric model which includes geometrical as well as non-geometrical data, has yet to be addressed comprehensively. The main focus of the work presented in this paper is on developing an approach to collecting non-geometrical data from distributed sources (including online and offline data sources) and semantically enriching the geometrical model generated from the PCD. The developed approach requires addressing several challenges in achieving such a comprehensive semantically rich model of an existing building from a PCD. In current practice, the non-geometric data is appended to the model manually by utilising commercial BIM software or stored in different file formats. Due to the commercial BIM software limitations indirectly capturing 3D models for existing buildings, some of the information that cannot be appended to the model is stored in different file formats, such as PDF, 2D paper-based CAD drawings, Excel spreadsheets etc. outside the model. One of the challenges involved in generating BIMs for existing assets is the management and manipulation of such data that is stored in different file formats [198]. In this work, an interim solution using CSV (Comma separated values) files has been used to manually capture all this information before converting them to RDF (Resource Description Framework) and ultimately to IFC (Industry Foundation Classes). This has been explained in detail in section on CSV to RDF and RDF to IFC Algorithms. However, there are other approaches for automatically capturing geometric as well non-geometric data and these have been discussed in section on Related Work. The reason CSV file has been used in this work is due to the lack of access to these other algorithms for proprietary reasons.

In recent years, several studies have been undertaken to make the parametric modelling process as effective and efficient as possible by developing various algorithms based on the PCD as the main data source. However, as mentioned above the generated models are 3D representations of building components that contain geometric data only. Although the mapped components are considered as parametric models, a semantically rich parametric model is still some way away. A semantically rich parametric model can be defined as a model that contains two types of data, viz. geometric data (e.g. coordinates (points), dimensions, element connectivity, etc.) as well as non-geometric data (e.g. material, colour, element constituents, load-bearing capacity, security rating, fire ratings, etc.) [042-333]. In contrast to the geometric data that can be extracted from PCD, the non-geometric data is not included in PCD and needs to be extracted from other data sources. In current practice, generally speaking, the non-geometric data is appended to the model manually by using commercial BIM software, or it is stored separately in various file formats. Interacting with the external data sources in different formats gives rise to the challenges around information exchange and interoperability.

To address this problem, a variety of schemas like ifcOWL [144] have been developed to make the distribution of the data on the web more effective and efficient. However, these schemas are not designed to generate building models. Instead, they are used to extract information for distribution from existing models. The ifcOWL schema is predominantly created from an existing IFC model through the process of converting IFC into OWL (Web Ontology Language) ontology by the implementation of IFC-to-RDF [014] and EXPRESS-to-OWL [144] algorithms. The process of developing such schemas mainly commences from an existing building model, which may or may not incorporate geometric and non-geometric data. On the other hand, at the time of writing this paper, such schemas are not supported by available BIM applications. Therefore, this research addresses the problem of generating IFC from RDF and is the reverse of what is currently possible to do as explained earlier.

In current practice, the proposed approaches for generating parametric models from PCD mainly focus on the identification and generation of geometries and shape primitives rather than the asset information (non-geometrical data) that is required in BIMs [009], and is crucial for O&M and other aspects of the BIM-based asset life-cycle information management process. On the other hand, the objective of schemas like ifcOWL is to store and share information on the web more effectively rather than generating 3D models. In light of all these issues, the proposed approach in this work uses semantic web technologies to capture information extracted from distributed online and offline sources prior to the generation of 3D models. This aggregated data can then be used to generate a comprehensive 3D model containing geometric as well as non-geometric data. Details of this proposed framework and its implementation are given in Sections 3 and 4.

1.1 Scope and Context

The project reported in this paper was instigated by a partnership between Historic Environment Scotland (HES) and the authors' institution. HES own 345 historic and heritage buildings in Scotland, and they launched an initiative to implement BIM for the operation and maintenance of their assets back in 2013. HES have developed a very comprehensive BIM strategy. The pilot project to implement their strategy was to develop a retrofit model of the Main Palace of Edinburgh Castle. The Main palace of Edinburgh Castle was built in the 16th century and consists of several unique features. HES laser-scanned the main palace and generated a 3D geometrical model of the main palace quite successfully. However, they hit a major bottleneck in converting the geometrical model into a more semantically rich parametric model, which would include relevant non-geometrical data required for O&M of the asset. The required non-geometrical data was scattered over several sources and was inevitably stored in different formats, including 2D drawing and PDF documents generated from online sources. More details about this project can be found in the 'BIM Pathfinder Projects' report carried out by the Scottish Future Trust (SFT) [295]. They, therefore, sought external assistance to address this challenge and hence, the study reported here started with a view to using their data to validate any proposed solution. The authors set out to develop a generic approach for all existing structures with further adaptations as required for unique assets like historical buildings.

2 Related work

In recent past, several studies have been conducted to develop automated or semi-automated approaches for generating parametric models, i.e. building geometries, by utilising PCD as one of the primary data sources and developing various algorithms to enhance the performance of the developed methods. These approaches are used to recognise building elements in the PCD for a variety of purposes like identifying discrepancies and similarities between as-designed and as-built models [306-053] as well as tracking the construction progress [075-089-238-341]. The general workflow of the proposed approaches can be classified into two processes, viz. 'Scan-vs-BIMs' focusing on the identification of correspondences and discrepancies between PCD (as-built) and the 3D model (as-designed), and 'Scan-to-BIMs' focusing on the generation of building geometries (parametric models) utilising PCDs directly. In addition, other approaches have been developed that use PCD to identify building elements to create pre-defined libraries of 2D and 3D building shapes [141-181].

2.1 Scan-to-BIMs

In contrast to 'Scan-vs-BIMs' process, an existing building may not have a 3D CAD model or indeed any model at all. In such cases, paper-based 2D drawings or digital documents are the only available information sources for generating BIMs. In this case, the procedure for generating BIM models is implemented through the 'Scan-to-BIMs' process by utilising the geometric data extracted from PCD as the primary data source. The data collected from an existing building is utilised to calculate, recognise, and detect building geometries. The approach proposed in Zhang et al. [079] focuses on the reconstruction of building elements in various real-world projects. Different data collection technologies, such as 3D laser scanning and Videogrammetry, is utilised to collect the data from existing buildings in the form of PCD. The main focus of this method is the identification of planar surfaces in the PCD due to the importance of planar patches in shaping 3D geometries and primitives [053-008-141]. A segmentation algorithm declared based on the unsupervised subspace technique [211] is utilised to retrieve linear relationships between elements in PCD. This technique is employed to identify the number of linear relationships, associated dimensions, and segmentation groups of points in PCD. The Maximum Likelihood Estimation Sample Consensus (MLESAC) [254] and Singular Value Decomposition (SVD) [255] methods are then applied to calculate and extract plane models from the PCD. The α-shape algorithm [255] is lastly used to extract the corresponding planar patches (surfaces) from the PCD as the final output of the proposed approach.

There are also other studies carried out that are based on 'Scan-to-BIMs' process capturing building geometries in the historical building modelling (HBIM) domain. The proposed approaches involved in historical buildings use geometric data extracted from the PCD to generate historical objects. As an example, the semi-automated approach proposed in Barazzetti [042] focuses on the identification of historical objects. Discontinuity lines that are extracted and calculated from the PCD are first reconstructed by using NURBs (Non-Uniform Rational B-splines) features. Reconstructed elements are then utilised to create surfaces. Subsequently, parametric models are then generated by the connection between identified surfaces. Similar approaches have been proposed in the heritage domain utilising PCD as the geometrical data source to develop automated or semi-automated algorithms for generating parametric models. Other examples of using the Scan-to-BIMs method for retrofit and historical buildings are studies undertaken by Banfi et al. [170] and López et al. [162].

2.2 Pre-defined libraries for parametric models

Methods for model generation based on pre-defined libraries focus on identifying building elements in PCD and creating libraries of parametric models based on the detected elements and project requirements. Building elements stored in the libraries are later used to generate parametric models. The semi-automated approach proposed in Dore and Murphy [141] is based on the development of a library of parametric components utilised for modelling architectural objects for historic buildings. This approach contains different rule-based algorithms which are developed to combine elements stored in a pre-defined library and to map building layouts in a heritage building environment. The detected elements in PCD are first compared to the objects in the pre-defined library, after which the matched candidate is used to generate the corresponding parametric model. More information about this approach can also be found in Murphy et al. [175]. The work presented in Apollonio et al. [116] focuses on generating profile-based libraries for HBIM. A library of shape profiles is first generated based on the architectural ontology for corresponding building elements. The pre-defined profiles are then used to generate the historical components in a BIM software. More examples of approaches that utilise pre-defined libraries for capturing 3D building objects can be found in Heidari et al. [123] and Brumana et al. [181].

2.3 Information management in the AEC industry

One of the main reasons behind BIM-driven project delivery in the AEC industry is the storage, sharing and reuse of information [230-165], between all stakeholders involved in projects [221]. BIM models generated in BIM applications are used as one of the information sources in BIM-enabled projects [302]. However, a BIM model generated in one BIM application may or may not be supported by other BIM platforms. Hence, open data exchange standards and schemas have been developed to enhance the communication between modelling applications. The IFC data model, as a data exchange standard in the building industry, is developed by buildingSMART over the past several years [165-312]. It is the most well-known and widely used set of standards for exchanging information about a building between diverse IFC-compliant BIM applications [303-305]. After almost two decades, the current version of IFC (IFC 4.2 at the time of writing this paper) has made considerable progress with more than 700 classes, thousands of attributes and a dense network of relationships between its classes [165-304]. However, it still has limitations for specific functionalities [014] like road and infrastructure [165-218] and indeed, historic and other existing assets [235]. In addition, it does not capture the full semantics needed for different aspects of a BIM process [221]. However, in spite of the limitations involved in IFC, it is still the de-facto and widely adopted standard for information exchange and interoperability between BIM-driven applications in the building industry [165-297-303].

Semantic Web technologies and standards, such as RDF, RDFS (RDF Schema), and OWL developed by W3C (World Wide Web Consortium) group, are also gaining popularity within the AEC industry. The use of Semantic Web technologies can be considered as alternative options for improving information interoperability in the AEC industry [217]. With regard to the AEC industry, these technologies are mainly used to store and share data about building projects on the web and indeed to improve the representation and distribution of data. The OWL ontology for IFC (also known as ifcOWL), for instance, has recently been proposed by Pauwels and Terkaj [144]. ifcOWL is basically generated from an existing IFC data model by translating IFC into OWL ontology through the implementation of the IFC-TO-RDF and EXPRESS-TO-OWL algorithms. The general concept behind the development of ifcOWL is to use Semantic Web technologies, such as RDF and OWL, to enhance the distribution of data, extensibility of the data model, data storage (on the web), consistency checking, and knowledge inference [144]. Another example of such schemas is the ifcJSON proposed by Afsari et al. [276]. The proposed schema, in general, focuses on using JavaScript Object Notation (JSON) serialization and converting the ifcXML format into an ifcJSON schema. This can subsequently be used for transferring data on the web and improving the interoperability of cloud-based BIM platforms.

2.4 Data fusion

Data fusion, also known as data integration, can be defined as the process of combining multiple data sources to improve the nature of information, which consequently improves the process of estimating the state of any entity in the integrated data [325-328]. In other words, the main idea behind the use of a data fusion framework is to integrate data collected from multiple data sources as well as the corresponding information gathered from associated databases for improving accuracies and specific inferences [344-309]. The fact is that individual and separated sources usually provide a portion of the data items needed for the analysis and decision making processes [329]. However, the use of an appropriate data fusion framework, in general, improves the reliability of data, the data detection and the data ambiguity reduction processes [328-309]. Hence, over the past decades, several studies have been undertaken to develop a generic and formalised data fusion framework in different domains, such as computer vision as well as defence and robotic. The Joint Directorate of Laboratory (JDL) fusion model [346], Omnibus model [345], and Dasarathy's fusion model [352] are examples of most famously developed and widely used data fusion models [329]. In terms of application, data fusion is not limited to a certain domain and can be applied to other domains like remote sensing and signal processing, monitoring purposes, offline and online textual data processing, construction, and engineering [325-326-348-351].

In terms of the use of data fusion models in the AEC industry, several studies have been carried out proposing data fusion-related frameworks for monitoring the progress of construction performance and activities [327], tracking on-site materials [328-330], and planning models for construction productivity [329]. The work presented in Pradhan et al. 2012 [329] proposes a planning approach for fusing data from multiple data sources to support the monitoring process of construction productivities based on Hierarchical Task Network (HTN) and GraphPlan methods. The HTN method as a domain-dependent algorithm is first used to generate an abstract plan for fusing multiple data sources. The existing GraphPlan algorithm as a domain-independent method is then utilised to generate data fusion-based plans for productivity-related queries. Another example is the process management framework for multisensory data fusion framework proposed in Shahi et al. 2015 [327]. This focuses on tracking the progress of construction activities, estimating the construction earned-value, and updating construction-related schedules. Different data types are utilised throughout the data fusion process. This includes sensory data sources like Radio-Frequency Identification (RFID), Ultra Wide Band (UWB), and PCD, as well as non-sensory data sources like progress, schedule, and inspection reports. Other examples for the use of data fusion frameworks in the construction industry can be found in the studies undertaken by Shahandashti et al. 2011 [328], Razavi et al. 2010 [330], and Soltani et al. 2018 [332].

2.5 Challenges and limitations

BIM models are considered as one of the essential parts of a BIM process, and they incorporate information that is crucial for procuring full advantage of BIM-enabled building projects. An appropriate parametric model that is fit for purpose for a BIM-based process of design, construction and O&M of assets should incorporate geometric and non-geometric data. The non-geometric data that is required in captured models is generally appended to the model either through a manual process or indeed eliminated from the model. Hence, several studies have been undertaken in the literature to improve the manual process of generating BIM models by developing automated or semi-automated approaches with varying success. These mainly use PCD as the primary geometrical data source to identify and recognise building elements. The final results of such approaches are simple shapes or primitives that only contain information about geometrical representation of building elements, such as coordinates, length, width, area etc. [043]. However, the non-geometric data needs to be appended to the model through a process by either converting 3D geometries into real BIM objects where the non-geometrical data can be attached to the model or creating new BIM objects based on the collected data and project specifications. Nevertheless, the PCDs do not include non-geometric data, thus resulting in parametric models that do not contain the critical non-geometrical data.

While some of the proposed approaches focus on improving existing information exchange standards and tools like IFC data model, others focus on developing new schemas, like ifcOWL [144] and ifcJSON [276] by utilising Semantic Web technologies and standards. The main idea behind the development of new schemas is to use existing information about a building and convert it into OWL ontologies, which are predominantly used to store and share the information on the web.

The developed schemas mainly focus on using integrated information exchange standards, predominantly IFC schema, and Semantic Web technologies to produce shareable data. The data used for implementing such schemas is extracted from an existing model. In fact, the model employed for creating shareable information may or may not incorporate required non-geometrical data. In current practice, information embedded in the model is extracted from it in the form of IFC by the use of BIM applications. This, in fact, signifies that if the IFC is extracted from a model that is generated based on the data presented in PCD, the extracted IFC still does not include non-geometric data. Subsequently, the shareable information will not incorporate this type of information. Moreover, at the time of writing this paper, the other limitation of these schemas is that available BIM applications do not support them as a source of information for generating BIM models. With regard to the identified challenges and limitations as well as research contribution, Figure 1 shows the general process of proposed approaches, developed schemas, and the framework proposed in this paper.

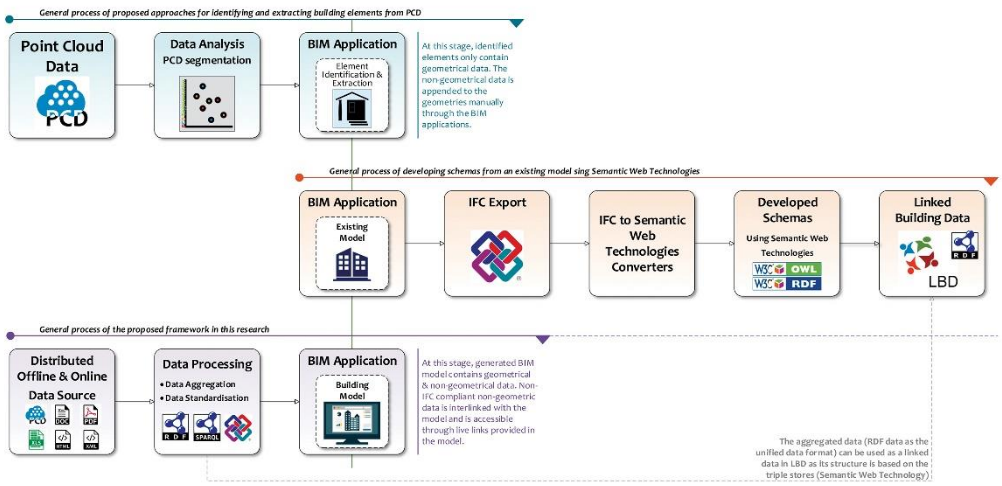

Figure 1: From top to bottom: The general process of generating building geometries using PCD as the primary data source, the general process of developing schemas using an existing model, and the general process of the framework proposed in this paper.

3 A framework for developing semantically rich 3D models

In current practice, the non-geometrical data that cannot be included with the model during the generation process of a BIM model is typically stored in diverse data formats, spread between offline and online data sources. The use of different data formats makes the process of data manipulation and management inefficient, and indeed difficult. Hence, a framework is proposed in this paper, which aims to address the challenges and limitations involved in generating semantically enriched BIM models from PCD. Figure 2 illustrates the proposed framework. The framework is composed of three main processes, viz. 1) Data Collection, 2) Data Processing, and 3) BIM Generation.

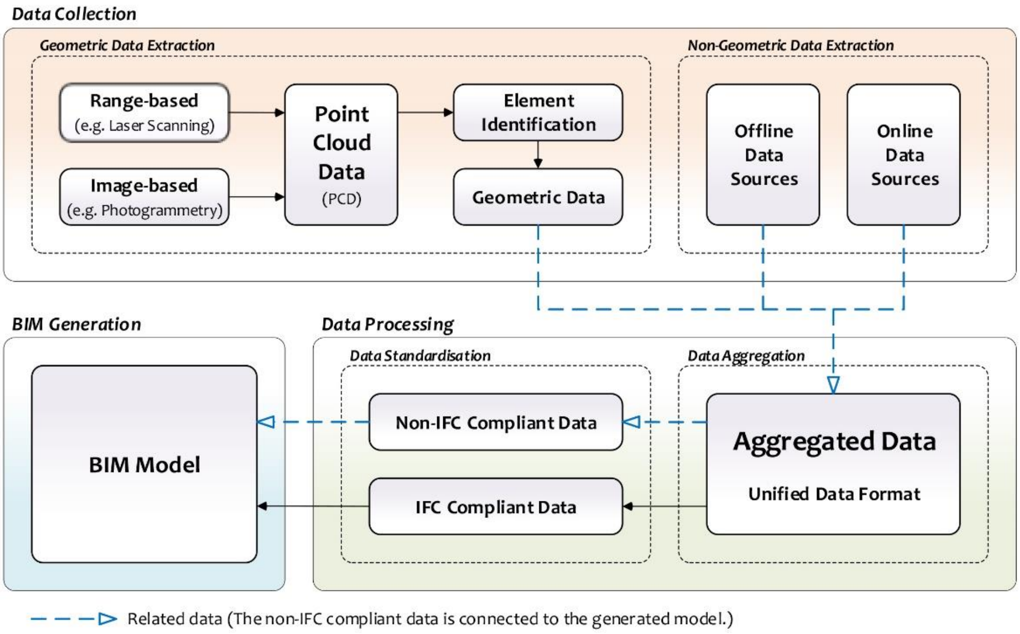

Figure 2: The proposed framework.

3.1 Data Collection

An appropriate parametric model that is fit for purpose for a BIM-based process of design, construction and O&M of assets should incorporate geometric and non-geometric data. Accordingly, the Data Collection process of the proposed framework consists of two sub-sections, including geometric data retrieval and non-geometric data extraction. The PCD is first utilised as the primary data source to retrieve the geometric data from identified building elements which only contain geometrical data. A variety of novel image-based and range-based data collection technologies have been utilised in the AEC industry and other domains like computer vision and robotic systems for collecting data from an environment or object. LiDAR (Light Detection And Ranging) also known as LaDAR (Laser Detection And Ranging) or 3D Laser Scanning as well as Photogrammetry and Videogrammetry are evaluated as a rapid, accurate, and commonly utilised solutions to collect data from existing and retrofit buildings within the AEC [009-042-310-334]. However, the use of each data collection technology individually comprises limitations. An example of these limitations is the Region Of Interest (ROI) of the laser scanner, which is a significant limitation in the data collection process. This also plays a significant role in enhancing the performance of data processing and BIM model generation processes [241]. Hence, a combination of different techniques, such as Laser Scanning and Photogrammetry, enhances the accuracy of the data which may affect the identification of building elements process directly [039]. For instance, the approach proposed in Klein et al. 2012 [015] utilises Laser Scanning and Photogrammetry technologies to collect the data and to enhance the process of reconstructing building geometries.

The output of the Laser Scanning technique is a set of point measurements, typically known as Point Cloud Data (PCD). In terms of utilising PCD as the primary data source for generating building elements, the output of Photogrammetry (images) and Videogrammetry (images extracted from video frames) are subsequently converted into PCD [075-322-323] by utilising relevant image processing algorithms and applications. The registered PCD can then be utilised to identify and extract geometrical elements. As an example of converting images into PCD, the work presented in Rashidi et al. 2015 [342] proposes an approach for generating PCD for outdoor and indoor settings through the process of converting photographs collected by a monocular camera into PCD. In current practice, the process of using PCD to identify and generate building elements is mainly carried out manually by using available CAD applications. However, semi-automated approaches have been proposed and developed over the past few years to make the element identification process more effective and efficient [093-089-306].

With regard to the developed approaches, the PCD segmentation process is carried out by declaring various algorithms based on the previously defined algorithms, such as unsupervised subspace learning technique, MLESAC, Singular Value Decomposition (SVD), α-shape, and shape grammar [265]. For example, the shape grammar procedure as a rule-based algorithm is used in Dore & Murphy [141] to design shape vocabulary, which is later used to create a library of parametric geometries. Elements in PCD are then matched to the objects defined in the library using ArchiCAD BIM platform. The work undertaken by Zhang et al. [079] is another example of reconstructing building elements from PCD. A combination of aforementioned algorithms (e.g., SVD and MLESAC) is utilised to extract planar patches from acquired PCD. The identified geometrical elements, such as lines, curves and surfaces, are then used to generate 3D building elements manually in CAD and BIM applications. As previously mentioned, the generated models only contain geometrical parameters, and the non-geometrical data is not included in identified elements. Notwithstanding, the geometric data like cartesian points (coordinates) and geometric properties like length, width and height can be extracted from the generated elements. This can be utilised as the geometrical data source in the proposed framework in this paper. The process of extracting geometrical data from PCD is shown in Figure 3. In addition to that, offline and online data sources are used to collect the non-geometrical data. In current practice, the non-geometric data is stored as offline and/or online data in different formats. These data sources are used to extract the required non-geometrical data presented in different data formats.

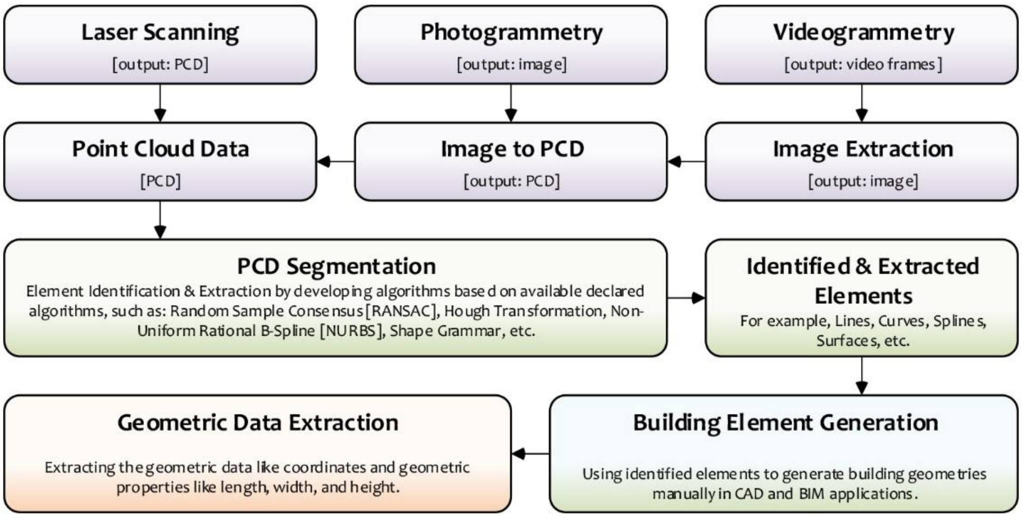

Figure 3: Data collection process: Geometric data extraction from PCD.

3.2 Data Processing & BIM Generation

The data processing step is composed of two sub-divisions, viz. data aggregation and data standardisation processes. The data collected from each data source at the previous step is composed of different formats, which makes the data manipulation and management process inefficient [329]. The use of different data formats also affects the sharing, long-term access, and preservation of the data. Hence, the use of a single standard format facilitates the data management process. In fact, it is well-known that the fragmentation within data sources in which data is stored in various formats makes the data correlation and management difficult and indeed ineffectual. In addition, data stored in different data formats cannot communicate in an effective and efficient manner which makes it difficult for the user to manipulate and manage it. However, a unified data format – data unified in a single standard – simplifies the process of data manipulation and management.

In the proposed framework, the collected data is first aggregated into RDF data as the unified data format. RDF was developed and agreed upon by World Wide Web Consortium (W3C) as a standard format for data interchange [262]. Being a framework for representing data, RDF, in general, can be defined as a method for describing data by defining the relationships between data objects, i.e., RDF describes data with the semantics embedded in it [155]. In addition, RDF is capable of integrating multiple data sources effectively [156]. Nevertheless, RDF is used in this study to facilitate the data analysis as well as the storage, share and reuse of the data [314-164]. At this stage of the proposed framework, the RDF data encompasses the geometric and non-geometric data. The data is then classified into two distinctive sub-divisions. The first includes data that is compliant with the IFC model (Ifc compliant data) and can be combined with the 3D model through the IFC format directly. In other words, with regard to the IFC structure, the geometrical data and a small portion of the non-geometrical data can be combined with the model through the process of IFC creation. On the other hand, the data – the latter sub-division of non-geometrical data (non-IFC compliant data) – that cannot be combined with the model through the IFC remain in the form of RDF data. This part of data is related, i.e. interlinked, with the generated 3D model. The data processing and BIM model generation processes of the framework are illustrated in Figure 4.

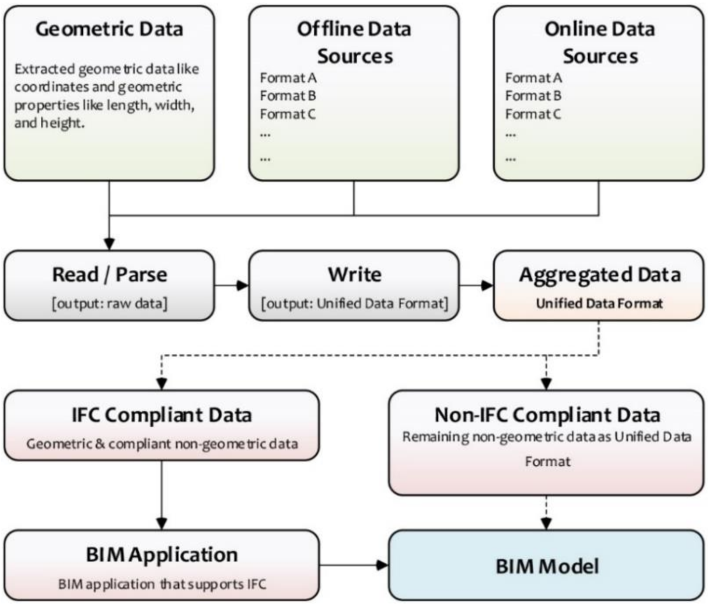

Figure 4: The process of data processing and BIM generation of the proposed framework.

4 HES BIM Projects: Edinburgh Castle

Edinburgh Castle BIM project carried out by Scottish Future Trust (SFT) and Historic Environment Scotland (HES) as a research study [261] for the Level 2 BIM implementation in Scotland is used in this study to identify challenges and limitations involved in generating parametric models for existing and retrofit assets as well as the management of large-scale data required to be embedded in models. As mentioned previously, HES is responsible for managing 345 Properties In Care (PICs), including Edinburgh Castle. The BIM model of Edinburgh Castle is generated manually based on the PCD collected from the main Palace Block (Figure 5). The Required Asset Information (RAI) is appended to the model manually through the process of either converting simplified geometries into family types in Autodesk Revit software and reloading created families into the model or creating new objects based on the available information. However, with regard to the HES project scope and objectives presented in Table 1, information that cannot be appended to the model through the aforementioned process is stored in various agreed data formats separately.

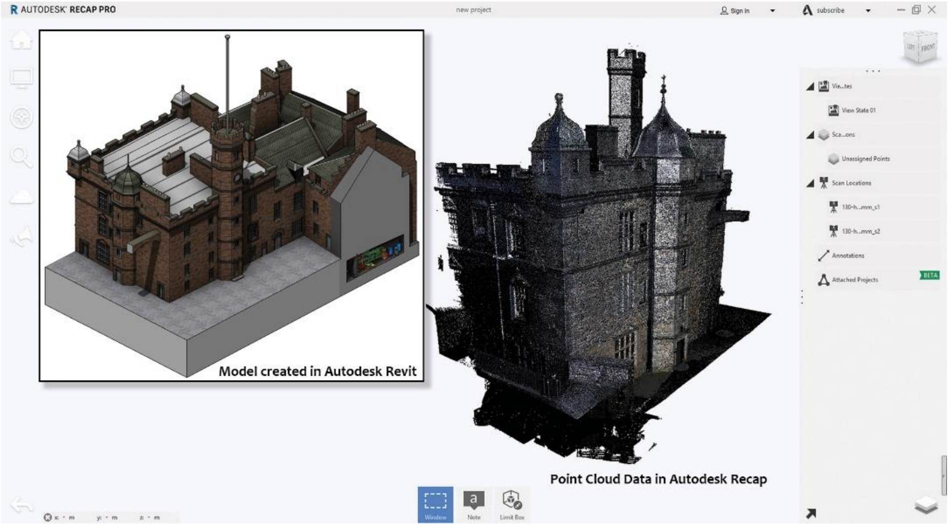

Figure 5: Edinburgh Castle: Main Palace PCD and the model (source: HES 2018).

Table 1: HES BIM Project Objectives and scope (source: HES 2017).

Project Objectives

1. Inform the development of a full business case setting out HES's BIM strategy, including an assessment of benefits, life-cycle costs and resource requirements, in order to secure senior management approval for the use of BIM as an integral component in HES's organisational processes.
2. Support the delivery of statutory obligations under the Scheme of Delegation between HES and the Scottish Ministers by contributing to the replacement of inefficient, ad hoc working methods with standardised and reliable information management and reporting processes.
3. Develop skills and knowledge of BIM tools and processes across all levels of HES whilst developing expert client competencies to manage the procurement of information from external supply chains in an effective manner.
4. Coordinate with and contribute to other ongoing HES Conservation Directorate information management work streams (please see Project Details section).
5. Engage with partners and stakeholders to contribute to the development of the Scottish BIM guidance and wider industry practices relating to the application of BIM to existing built assets.
6. Improve access to high-quality asset information in order to improve the quality of decision-making and minimise the likelihood of abortive work, additional costs, disputes and potential reputational damage arising from the use of uncoordinated or unreliable information.
7. Future proof HES and enhance its reputation for using cutting-edge digital tools to care for and manage the historic environment.

Project Scope

1. As-existing Asset Information Model (AIM) of the Palace Block in line with the identified Asset Information Requirements, consisting of:
   - 3D geometric and analytical models (architectural, structural and services) reflecting the existing physical conditions of the Palace Block. The models will be generated principally on the basis of laser scan point cloud data, supplemented by other information sources (e.g. legacy drawings) as required.
   - Asset attributes information to support the identified usage of the AIM.
2. Appropriate data structures, templates and standards.
3. Outputs in agreed formats.
4. Analytical post-project evaluation to inform wider organisational engagement with BIM going forward.

Hence, a primary asset information requirements model, also known as Engine Shed Asset Information Model (AIM), has been created by HES representing the information that is required to be included in BIM models. This is subdivided into seven categories, viz. Identity, Spatial, Architectural & Structural, Electrical, Mechanical, Environmental, and Operational classifications, which includes both geometrical and non-geometrical data. Figure 6 illustrates the distribution of a portion of the required information, in particular, Identity, Spatial, Architectural, and Operational data, which are needed to be embedded in generated building components. The Spatial section represents the geometrical data which can be extracted and calculated from PCD. This part is considered as the IFC compliant data. Other sections represent the non-geometrical data that can be extracted from distributed online and offline data sources. These are considered as the non-IFC compliant data. The AIM is utilised in this research to structuring the CSV files based on the HES BIM project specifications. An example of using AIM classifications to create CSV data manually for the project and wall object entities is illustrated in Table 2.

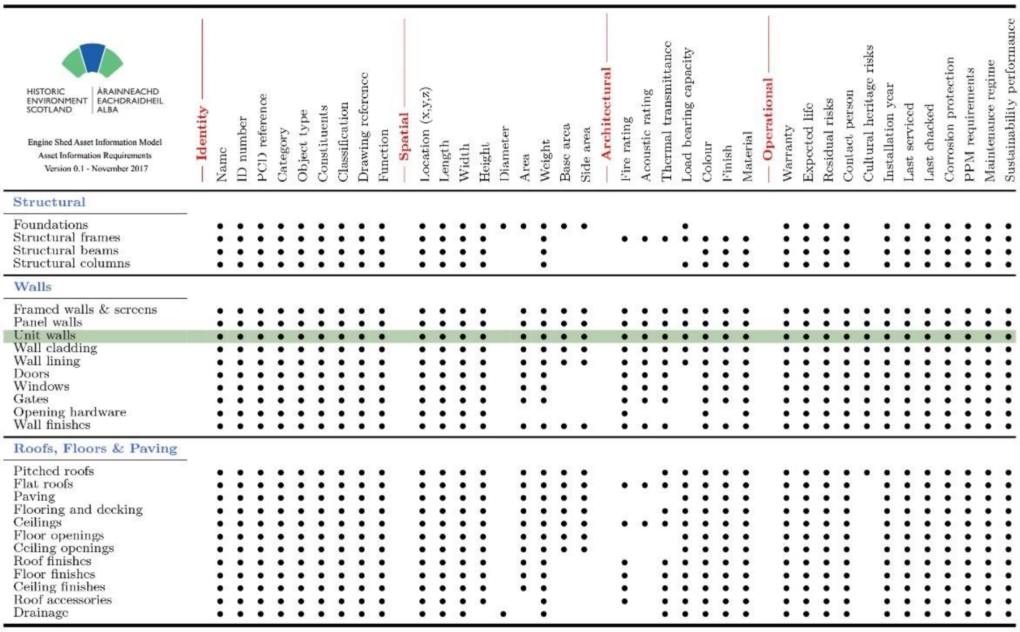

Figure 6: A portion of HES BIM Project AIM (Source: HES 2017).

5 Implementation of the proposed framework

As mentioned earlier, the proposed framework consists of three processes, viz. data collection, data processing and BIM generation. These are implemented through three key steps, viz. 1) the creation of CSV files representing the geometric and non-geometric data retrieved from PCD, offline and online sources, 2) the conversion of CSV files into RDF data, and 3) the translation of RDF data into IFC. The data collection process includes two distinctive data types, viz. geometric and non-geometric data. Different data sources -- PCD, offline and online -- are utilized to collect the data. While the geometric data can be retrieved directly from the elements extracted from PCD, the non-geometric data can be extracted from diverse data formats stored as offline and online data. As previously mentioned, this research does not focus on identifying and extracting building elements from PCD as this part of the framework is widely covered by other studies in the literature. Hence, the geometric data, such as coordinates, length, width and height, is extracted from the model generated in BIM applications in accordance with the acquired PCD.

Table 2: The distribution of Project and Wall entities data based on AIM.

Entity	Identity	Spatial	Architectural	Operational
Project	number category description name			
			phase	
				dateTime authorName authorEmail authorCompanyName
			fileSchema	
				projectOrganization purpose fullDescription userPurpose
	internationalLocation companyAddressLine companyPostalBox companyTown companyRegion companyPostalCode companyCountry			
		originCoordinateX originCoordinateY originCoordinateZ trueNorthX trueNorthY		
			representationMode	
		spatialDimensions		
			scale	
Wall Objects	pcdReference tagNumber			
			connectedTo	
	name id			
				editionyear
	category function objectType constituents classification drawingReference			
				sustainabilityPerformance codePerformance
			material colourR colourG colourB	
				imageRef contactPerson lastServicedYear risks
			fireRating acousticRating loadBearingCapacity	
				hesLocation
			openingDoor openingWindow wallDirection	
		locationCoordinateX locationCoordinateY locationCoordinateZ length width height baseArea sideArea volume weight		

First, the Comma Separated Values (CSV) format is utilised to gather geometric and non-geometric data in one place. The CSV file format is used to represent the geometric data that can be retrieved from elements captured from PCD as well as the non-geometric data that is stored in different data formats, such as 2D drawings and documents. A CSV file is a simple delimited text file in which values are separated by commas and stored in a tabular format as plain text. CSV files are manually created for different parts of the project, like the site, building, and building elements. Each CSV file contains data about a particular category of building elements. The CSV files are stored in a repository and used as the input data in the next step representing geometrical and non-geometrical data related to each building element.

The next step is to aggregate the data into a unified data format. An appropriate unified data format should facilitate the storage, share and reuse of data over the long term. The Resource Description Framework (RDF) – as a Semantic Web standard and technology – is utilised as an open standard format to structure the unified data format from previously-defined and -stored CSV files which represent the geometrical and non-geometrical data. The process of converting the data presented in the CSV files into RDF is carried out automatically through the procedure shown in Figure 7. As mentioned previously, the structure of CSV files, as well as the RDF data, are designed based on the HES primary asset information requirements model (Engine Shed AIM).

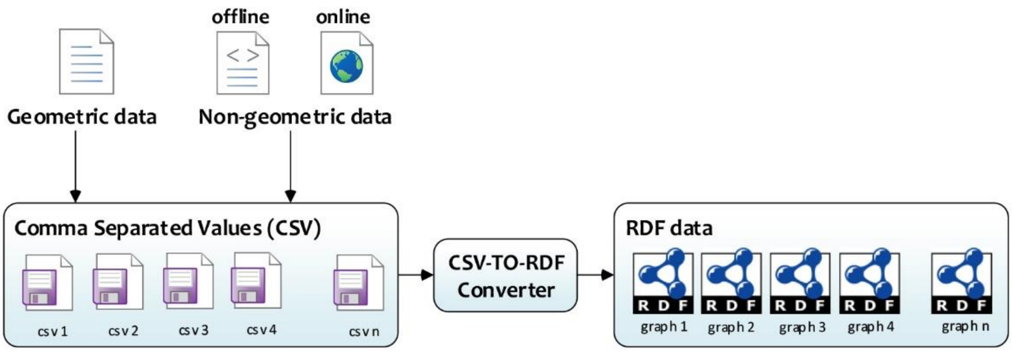

Figure 7: The process of converting *.csv files into RDF data.

The generated RDF data is utilised as the input data in the RDF to IFC translation process. As mentioned before, the aggregated data represented in the form of RDF is classified into two sub-sections, i.e. 'IFC Compliant' and 'Non-IFC Compliant' data. The first section includes data that can be combined with the model through the process of the IFC creation. This includes geometrical data and a portion of the non-geometrical data that is supported by IFC schema. However, the remaining non-geometrical data that cannot be appended to the model by IFC remains in the form of RDF data. The first sub-section of the RDF data (IFC Compliant Data) is translated into IFC through the process shown in Figure 8. This procedure generates a single IFC file by using the data presented in RDF graphs. The IFC file thus generated is then used to generate the BIM model by importing the IFC file into any BIM platform that supports this format. However, the remaining non-geometrical data is linked to the model. This data is shared on the web and linked to the model through the corresponding links added to the model properties through the IFC generation process.

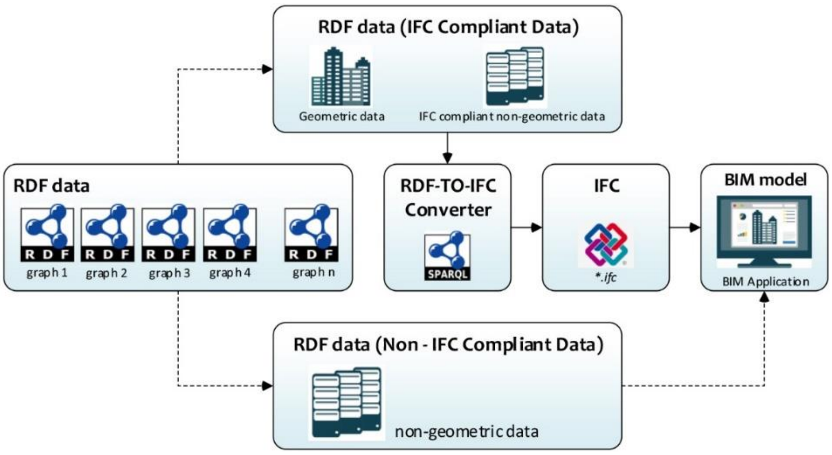

Figure 8: The process of converting RDF into IFC and generating the BIM model.

5.1 CSV-TO-RDF & RDF-TO-IFC Algorithms

The process of converting CSV files into RDF is achieved through the implementation of a CSV-TO-RDF algorithm (Algorithm 1). As shown in Algorithm 1, CSV files are used as the input data, and the Turtle serialisation of RDF is the output data of this algorithm. The representation of RDF data is based on simple statements, also known as triples, consisting of subjects (instances), predicates (properties), and objects (values). Where Subjects and predicates are declared as URIs that behave as unique identifiers, and objects can be declared either as URIs or Literals [155-156]. The algorithm iterates through CSV files as well as the data presented in individual CSV files. It then generates RDF data related to each building element. Triples (statements) generated based on the information extracted from CSV data is then stored as individual Turtle models.

```
Algorithm 1: CSV − TO − RDF Algorithm
  Input: *.csv
  Output: *.ttl
procedure CSV − TO − RDF(*.csv)
foreach csv ∈ csvDir do
    if (csvDir == ∅) then
        continue
    headerData ⟵ csvHeader.Split()
    define columnData List<string[ ]>
    while csvRow == csv.NextRow() do
        columnData.add(csvRow.Split())
    declare stringCells[headerData][columnData]
    if (columnData ≠ ∅) then
        continue
    declare turtleModel StringBuilder
    declare and set fileTitle String
    declare and set turtlePrefixes String
    turtleModel.append(fileTitle, turtlePrefixes)
    foreach headerInstance ∈ headerData do
        foreach cellValue ∈ columnData do
            declare and set turtleSubject
            declare and set turtlePredicate
            declare and set turtleObject
            set turtleModel
            turtleModel.append(turtleSubject, turtlePredicate,
              turtleObject)
    write turtleModel.ttl
    rdfDir.append(turtleModel.ttl)
next
```

Algorithm 1: CSV-to-RDF Algorithm.

IFC data model can be presented in the form of various formats, such as IFCXML (*.ifcxml) and IFC STEP (*.ifc). The IFC4 version of the STEP format is employed in this research to standardise the IFC compliant data presented in the form of RDF. IFC, in general, is structured based on two main parts, including the HEADER and DATA sections. The information about the file is presented in the HEADER section, and the project-related information, i.e. the information about building entities in a project, is presented in the DATA section. The process of translating RDF data into IFC is implemented through an RDF-TO-IFC algorithm (Algorithm 2). As shown in the algorithm, RDF data is used as the input data, and the output of this algorithm is a single *.ifc file.

Algorithm 2: *RDF – TO – IFC* Algorithm

```
Input: *.ttl
Output: *.ifc
procedure RDF – TO – IFC(*.ttl)
declare ifc4DataModel StringBuilder
foreach (turtleModel.ttl ∈ rdfDir ) do
    initialise iso1030321StepDef (ISO-10303-21;)
    declare headerSection (HEADER;)
    set headerEntities (fileDescription, fileName, fileSchema)
    end headerSection (ENDSEC;)
    initialise dataSection (DATA;)
    declare and set primitiveEntities
    ifc4DataModel.append (iso1030321StepDef (ISO-10303-21;), headerSection (HEADER;),
      headerEntities, headerSection (ENDSEC;), dataSection (DATA;), primitiveEntities)
    if (rdfDir ≠ ∅) then
        foreach (wallData.ttl ∈ rdfDir) do
            if (geometryRelEntities = ∅) then
                generate geometryRelEntities
                generate wallRelEntities
                ifc4DataModel.append (geometryRelEntities, wallRelEntities)
            else
                generate wallRelEntities
                ifc4DataModel.append (wallRelEntities)
end dataSection (ENDSEC;)
end iso1030321StepDef (END-ISO-10303-21;)
ifc4DataModel.append (dataSection (ENDSEC;), iso1030321StepDef (END-ISO-10303-21;))
write ifc4DataModel to *.ifc file
save *.ifc file
end procedure
```

Algorithm 2: RDF-TO-IFC Algorithm.

The algorithm generates the IFC entities, including HEADER and DATA entities, by iterating through the turtle models generated in the previous step (the output of CSV-TO-RDF algorithm). Depending on the data required for each IFC entity, algorithm extracts relating data from turtle models. The translation process is implemented through two general sections. The first section includes information that is associated with the common project information, such as project units, application, directions, 2D & 3D origin coordinates, and axis placement, which is later used to represent elements (primitiveEntities in the algorithm). The latter section includes entities that are related to the geometrical representation of objects in the IFC data model, which can be assigned to one or more objects according to the IFC specifications (geometryRelEntities in the algorithm). These are produced once during the generation of the first object. This section also includes entities that present each object (wallRelEntities for wall objects in the algorithm). The algorithm then writes and saves the IFC data model into a single IFC file. This is then used to generate the model by importing the IFC file into any BIM-driven platform that supports this format.

Application wise, there are several APIs available for generating RDF and IFC data, such as the Apache Jena API for generating RDF graphs and the IFC Java ToolBox for creating IFC files. However, this research uses its own code to generate RDF and IFC data. These are first created as strings and array values and then written into the corresponding file format, i.e. turtle models into *.ttl file format and IFC data model into *.ifc file format. In addition to that, Java programming language is used to implement the conversion of CSV files into RDF data (CSV-TO-RDF algorithm) as well as the translation of RDF data into IFC (RDF-TO-IFC algorithm) processes. This includes 122 classes, including functions and methods, and approximately 10000 lines of code for two algorithms.

6 Example Applications

The proposed framework described in the previous sections is applied to a prototype consisting of two wall objects. The information about different aspects of the prototype is first recorded and stored as CSV files. This includes geometrical and non-geometrical data about the project, site, building, building storeys, and building elements. The CSV files are then employed as the input data to generate RDF data. The generated models are then utilised as the input data for translating RDF into IFC. The result of the implementation of described processes is a single IFC file (*.ifc) which contains data about two wall objects. The created IFC file is then employed to generate the BIM model by importing the IFC file into Revit BIM application. Figure 9 shows the generated model. In addition, generated objects function as BIM objects and their type of specifications can be modified in BIM software directly (Figure 10).

The geometrical data and a portion of non-geometrical data that can be combined with the models are included in the created IFC file and presented as IFC parameters in the BIM software. However, the non-IFC compliant data, predominantly non-geometrical data that cannot be presented by IFC, is interlinked to the model as RDF data. As mentioned previously, the RDF structure consists of three parts, also known as triples which construct a statement including a subject, predicate, and an object. The subject and predicate are declared as URIs, and the object can be declared either as a URI or literal value. The subject URIs are the links to the entities that are provided in the model. The non-IFC compliant data can be accessed through these links by importing the IFC file into any BIM platform that supports this format or opening the model generated from the IFC file in BIM applications like Revit, BIM 360, and A 360 platforms. In addition to that, these links are included in the model during the process of translating RDF into IFC based on the IFC entity specifications (e.g. IfcPropertySingleValue and/or IfcURIReference entities). As shown in Figure11, Autodesk A360 is used to access the model and its properties as well as the aforementioned links provided in the model.

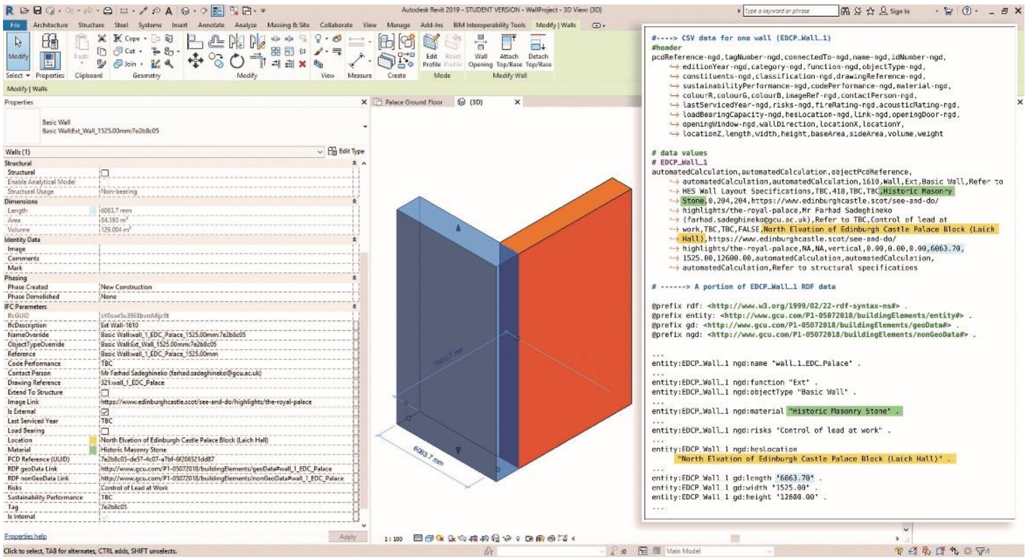

Figure 9: The model generated in Revit software using the IFC file.

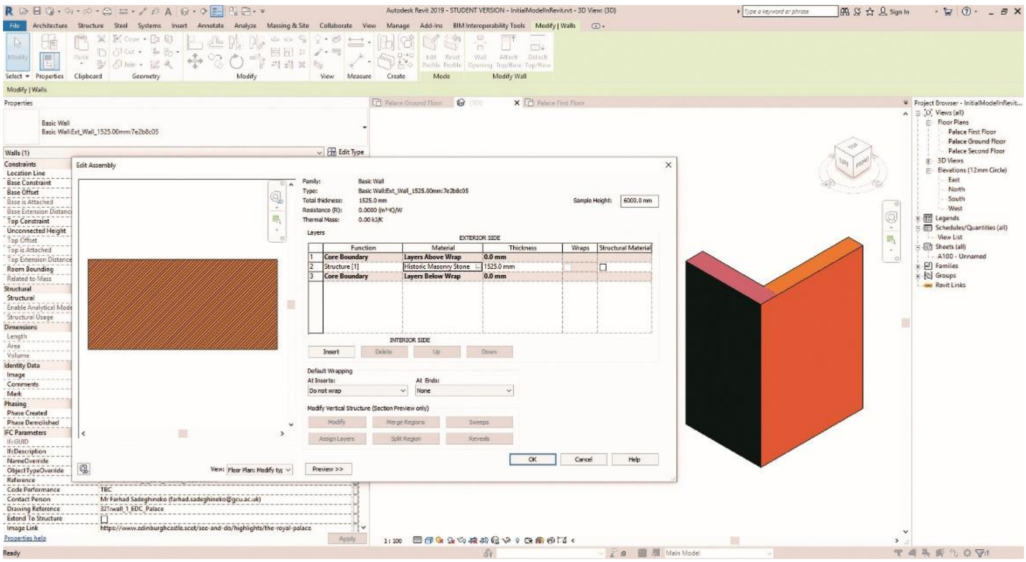

Figure 10: The properties of generated objects can directly be modified in the application.

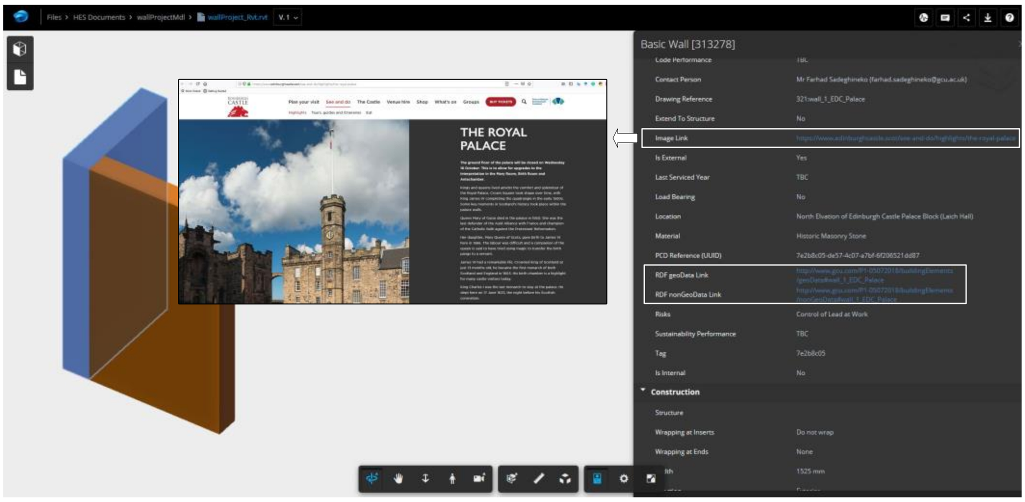

Figure 11: Accessible links for correlated data provided in the model.

As mentioned previously, in terms of the performance flexibility of the proposed framework, it is not limited to a specific building type and can be applied to any type of building, including new, existing and retrofit assets. The previously described algorithms used to implement the framework and to generate the previous example is also used in the following more complex example. This includes multiple wall components and other building objects distributed in two distinctly different floor plans of an existing building. The same approach is used to create the CSV data for different parts of the building, such as the project, site, building, building storeys, external & internal walls, and slabs. The RDF data is then generated for individual elements. The final output of this example is a single IFC file containing 1490 lines of data. The following figures (Figure 12, 13, and 14) illustrate the model generated in the Revit platform. The generated *.ifc file can be used in any BIM-driven application that supports this format. Accordingly, Figure 15 shows the generated model in BIM 360 environment. As shown in Figure 15, the subject URIs included in the model are live links (Linked RDF data) and can be used to access the additional data that cannot be combined with the model through the process of creating the IFC file.

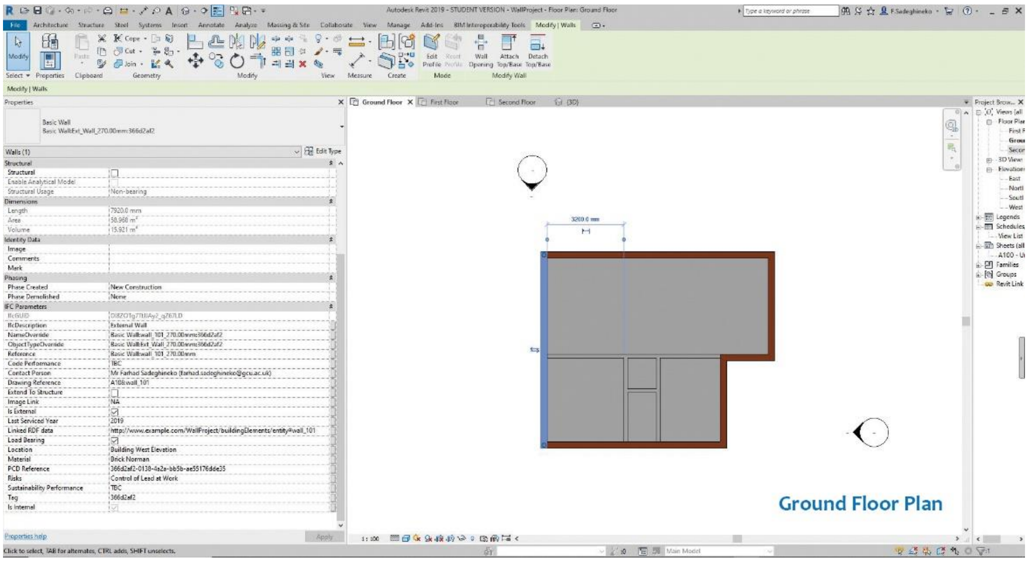

Figure 1: The ground floor plan view of the generated model.

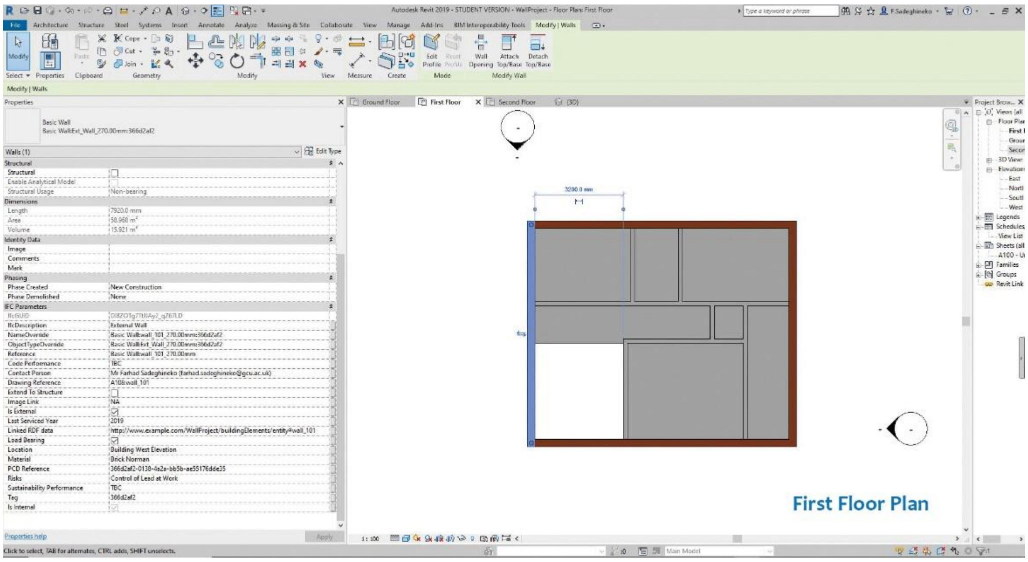

Figure 2: The first floor plan view of the generated model.

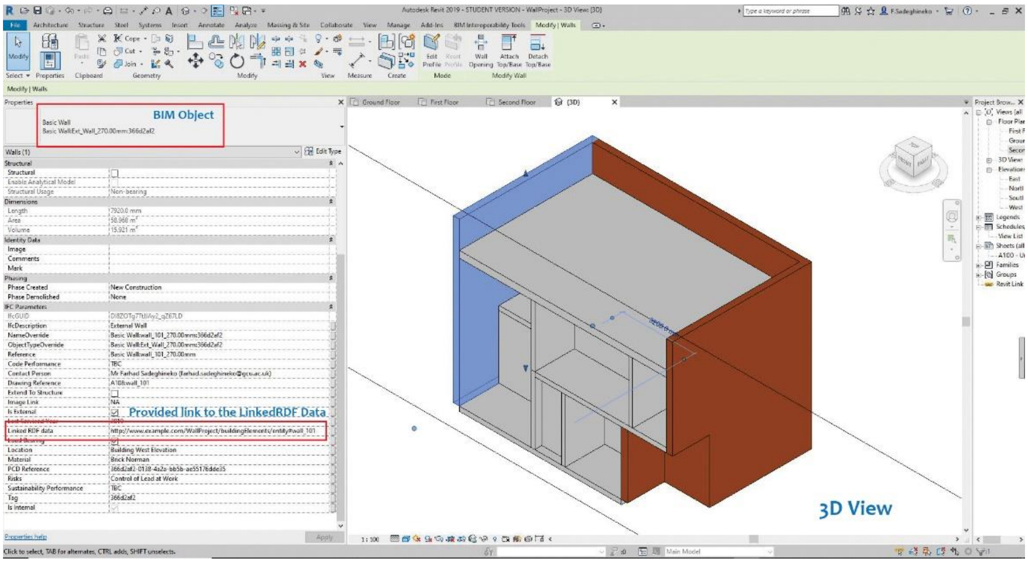

Figure 3: 3D view of the generated model.

7 Discussions and limitations

While some of the existing approaches for identifying and extracting building elements from PCD focus on generating building components that only incorporate geometric data, other approaches use schemas like ifcOWL that focus on creating shareable data which are mainly used to store, share and reuse data on the web. The latter group of approaches mainly use existing data, predominately IFC that is extracted from a previously generated building model which may or may not include all the required data. In addition, such schemas are not capable of generating building models.

The proposed framework in this study can be seen as a solution to the challenges and limitations involved in generating semantically enriched parametric models from PCD, which include geometrical as well as non-geometrical data.

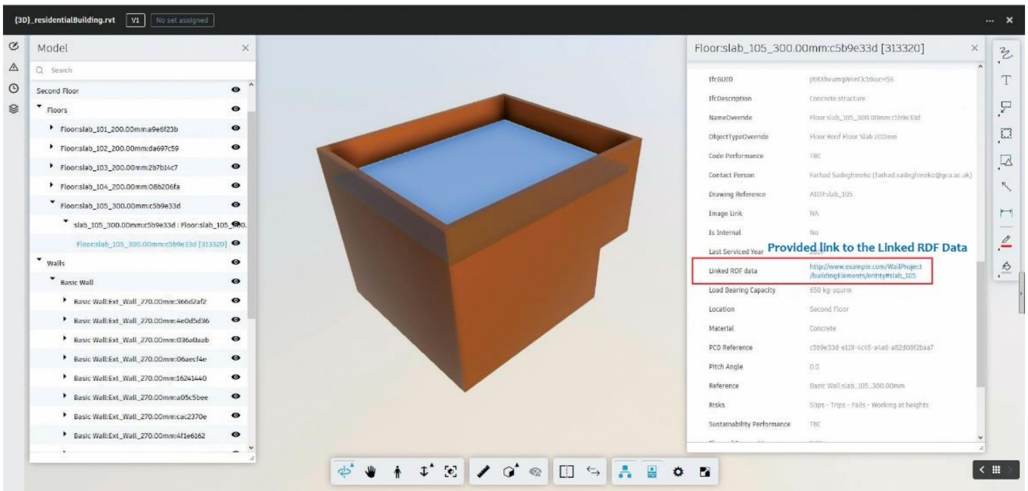

Figure 4: The model opened in BIM 360 environment.

The examples presented in this paper validate the potential of the framework. However, the following points must be taken into account and further improvements made in order to develop a more effective and efficient automated process for generating BIM models for existing assets:

- The incorporation of algorithms developed for identifying and extracting building components from PCD, which results in the extraction of geometric data required for generating the initial 3D shapes.
- As previously mentioned (see Data fusion section), several frameworks are developed and proposed in order to integrate data collected from diverse data sources. Hence, the use of appropriate algorithm(s) for extracting non-geometrical data from different offline and online data sources will improve the data aggregation process. This will eliminates the use of CSV files which is a limitation of the proposed framework as they are created manually.

8 Conclusions

A framework for developing parametric models for a BIM-based process of design, construction and O&M of assets should incorporate geometric as well as non-geometric data. Current approaches that focus on generating 3D model by using PCD as the main geometrical data source, mainly centre around identifying geometries in PCDs rather than on any other information required to be embedded in 3D models. In current practice, the non-geometrical data is appended to the model manually by utilising commercial BIM software or stored in different data formats separate from the model. The use of different data sources makes the process of data manipulation and management ineffective, and indeed error-prone due to human intervention. However, a unified data format – data unified in a single standard format – simplifies the process of data manipulation and management. On the other hand, a variety of schemas like ifcOWL have been developed to distribute data on the web efficiently. However, these are not designed to generate building models. Instead, they are used to extract information from existing IFC-compliant models for data distribution purposes.

The framework proposed in this paper aims to address the challenges and limitations involved in generating semantically rich 3D models from PCD. The framework consists of three distinct processes of Data Collection, Data Processing, and BIM Generation.

These are implemented through three key steps, viz. 1) the creation of CSV files representing the geometric and non-geometric data that can be retrieved from PCD, offline and online sources, 2) CSV to RDF conversion, and 3) the RDF to IFC translation. The RDF data is utilised as the unified data format to aggregate the geometrical and non-geometrical data. IFC is the most popular and widely used set of standards for exchanging information about a building between diverse IFC-compliant BIM applications. This format is utilised to translate the IFC-compliant data present in RDF data into IFC. The IFC file, thus created, is subsequently used to generate the BIM model by importing the file into any BIM application that supports IFC format. However, the non-IFC compliant data that cannot be combined with the model remains in the form of RDF data which is related (interlinked) to the generated BIM model by ………..

The use of RDF as a unified data format facilitates data management, in particular, large-scale data, i.e., it simplifies the data storage, sharing and reuse. In addition, being a widely tested semantic web technology and standard for data modelling, RDF is capable of representing high-quality connected data and provides the foundation for publishing and linking data. Hence, the use of RDF in the proposed framework facilitates data merging and linking. In other words, the geometric and non-geometric data presented in the form of RDF can be linked to other corresponding data sources if required. Having a uniform structure consisting of three linked data pieces (triples), the use of RDF also provides a standardised approach for interlinking and accessing the data in a formal and machine-processable manner. The other advantage of using RDF is that its use reduces the scale of the data by sharing equivalent data between similar components in a project which can also be employed in other projects if required.

The framework presented in this work is a semi-automated process where the collected data – geometric and non-geometric data that can be retrieved from PCD, offline and online data sources – is represented in the form of CSV format manually. However, the process beyond this point for creating RDF as well as the IFC is automated. The results are promising, and the future work of this study is to generate a fully automated process by eliminating the use of csv data, which are created manually and require human intervention, and using PCD data directly.

Acknowledgement

Authors would like to express their gratitude to Dr Vajira Premadasa of Historic Environment Scotland (HES) for providing assistance and support in the research reported in this paper.

References